BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

The Patent Office

PCT/GB 00 / 0 1 0 1 5

17 MARCH 2000

INVESTOR IN PEOPLE

GB00 / 1015

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

09/937100

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

REC'D 24 MAY 2000
WIPO PCT

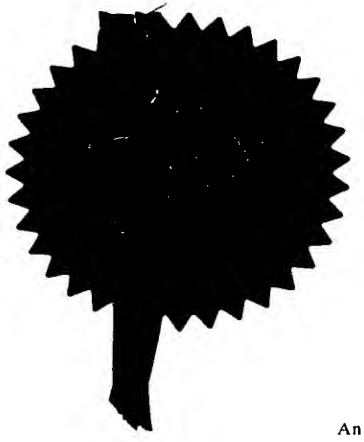
I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

I also certify that the attached copy of the request for grant of a Patent (Form 1/77) bears an amendment, effected by this office, following a request by the applicant and agreed to by the Comptroller-General.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.



Signed

Dated 31 March 2000

Patents Form 1/77

1977
(Rule 16)

The Patent Office

THE PATENT OFFICE
A
31 AUG 1999
RULE 97
RECEIVED BY POST

1/77

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

9920503.1

2. Patent application number
(The Patent Office will fill in this part)

31 AUG 1999

3. Full name, address and postcode of the or of each applicant *(underline all surnames)*

BIOVATION LIMITED
CROMBIE LODGE
ABERDEEN SCIENCE PARK
BALGOWNIE DRIVE
ABERDEEN AB22 8GU

Patents ADP number *(if you know it)*

7527536001

If the applicant is a corporate body, give the country/state of its incorporation

INCORP. IN SCOTLAND

4. Title of the invention

PROTEIN ANALYSIS

5. Name of your agent *(if you have one)*

"Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)*

AS (3)

Kilburn & Strode
20 Red Lion Street
LONDON
WC1R 4PJ

Form 5.177 a.300

Patents ADP number *(if you know it)*

12500

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*
—		

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*
—	

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*

a) any applicant named in part 3 is not an inventor, or
b) there is an inventor who is not named as an applicant, or
c) any named applicant is a corporate body.
See note (d))

YES

9. Enter the number of sheets for the following items you are filing with this form. Do not count copies of the same document.

Continuation sheets of this form	—
Description	18
Claim(s)	—
Abstract	—
Drawing(s)	—

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature

Date 30/8/99

12. Name and daytime telephone number of person to contact in the United Kingdom

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*

b) *Write your answers in capital letters using black ink or you may type them.*

c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*

d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*

e) *Once you have filled in the form you must remember to sign and date it.*

f) *For details of the fee and ways to pay please contact the Patent Office.*

PROTEIN ANALYSIS

The present invention relates to methods for analysing mixtures of proteins. In particular, the invention relates to methods to compare proteins between different cells and tissues. The invention involves the combination of digestion or cleavage of protein mixtures, and subsequent analysis of mass. The invention also preferably involves the fractionation of proteins or peptide fragments.

Current methods to analyse *en masse* complex mixtures of proteins such as in mammalian cells or tissues require that the proteins are separated by technologies such as two dimensional (2D) gel electrophoresis. For this technology, cellular proteins are usually separated on the basis of charge in one dimension and on the basis of size in the other dimension. Proteins can either be identified with reference to the electrophoresis migration pattern of a known protein or by elution of the protein from the electrophoretically separated spot and analysis by methods such as mass spectrometry and nuclear magnetic resonance. However, limitations of the 2D protein gel method include the limited resolution and detection of proteins from a cell (typically only 5000 cellular proteins are clearly detected), the limitation to identification of separated proteins (for example, mass spectrometry usually requires 100fmoles or more of protein for identification), the specialist nature of the technique and the difficulty in automating the technique in order to achieve very high protein analysis throughputs. There is thus a need for superior methods to analyse complex mixtures of proteins *en masse* especially using methods without gel electrophoresis and methods which are easy to automate.

The core of the present invention is that proteins are either digested or cleaved into smaller peptide fragments and then subjected to mass analysis especially by mass spectroscopy. Preferably, there will also be one or more protein or peptide fractionation steps to limit the complexity of the protein or peptide mixture being subject to measurement of mass analysis typically as mass-to-charge ratio measured by mass spectroscopy. Optionally, proteins or peptide fragments may also be conjugated with a "chemical tag" to assist in fractionation.

The major aspect of the invention provides for cleavage of proteins using proteases or chemical methods, fractionation of the peptide mixture thereby produced and subsequent mass analysis. One preferred method for fractionation of peptides is by using affinity reagents such as antibodies or solid phases or reactive chemical groups to isolate specific peptides or mixtures of peptides for subsequent mass analysis. Affinity reagents such as monoclonal or polyclonal antibody preparations can be used to retrieve individual peptides or sets of peptides from the peptide mixture for subsequent mass analysis. Alternatively or additionally, affinity reagents can be used to eliminate peptides from the mixture whereby the mixture is itself subsequently subjected to mass analysis. The affinity reagents can either bind by virtue of specific sequences or structures in peptides or by virtue of specific chemical groups either as natural constituents of the peptides or as chemical tags which are added to the peptides either before or after cleavage.

For analysis of larger mixtures of peptides, panels of mixed antibodies such as those provided by recombinant libraries of antibody variable region fragments (including single-chain antibodies) can be used in order to isolate subsets of peptides for subsequent analysis. Such panels of monoclonal antibodies will include a wide range of peptide specificities which could be achieved, for example, by pre-absorbing antibody libraries on the peptide samples of interest or by immunising animals with peptide samples of interest and collecting polyclonal antisera or generating panels of monoclonal antibodies. Then individual or mixtures of the selected antibodies are used to isolate (or eliminate) the specific subsets of peptides from a test sample. Subsequent mass analysis of a range of peptides can facilitate the detection of differences in specific proteins between test samples.

Fractionation of peptides can be achieved using affinity reagents other than antibodies. Generation of antibodies to all peptides in a mixture is difficult and is highly dependant on the number of peptides in a mixture and the facility for individual peptides to be bound with reasonable affinity to antibodies ("antigenicity"). With a very large peptide mixture, a limitation is redundancy whereby antibodies with the same peptide specificities are repeatedly represented whilst antibodies to other peptide specificities are underrepresented or absent. This may cause a particular protein to not be mass analysed if none of the peptides from a particular protein are bound by an antibody. Therefore, a particularly useful method is to isolate N or C terminal peptides (or both) from a protein by preabsorption of the protein to a solid phase via its N and/or C terminus prior to cleavage or by chemical tagging of the N and/or C terminus for subsequent isolation after cleavage. In principle, this then should lead to recovery of all N and/or C terminus peptides representing all proteins from the sample. Such isolation of N and/or C terminal peptides is greatly facilitated by the differential reactive nature of the N terminal amino group and the C terminal carboxyl group in the protein compared to internal amino and carboxyl groups. As an additional step, such isolated N and/or C terminal peptides can then be fractionated further prior to mass analysis using other affinity reagents which either recognise specific peptide sequences or which recognise chemical tags on the peptides. The invention also allows for sequential conjugation of different chemical tags to the protein / peptide mixture especially where N or C termini are sequentially exposed by specific cleavage of the protein / peptide and whereby the N or C termini (or both) are conjugated with a specific chemical tag upon exposure of that termini. This aspect of the invention therefore provides for a series of protein fractions with a range of conjugated chemical tags introduced at the termini, such fractions being isolated using an affinity reagent which binds to the tag. As an alternative to a chemical tag at the terminus of the protein molecule, chemical tags can also specifically be attached to non-terminus amino acids such that internal peptides can be isolated via an internal chemical tag. Unique chemistries are available for attachment of ligands to several specific amino acids, for example to the ϵ-amino groups of lysines, the thiol groups of cysteines and the carboxyl groups of aspartic and glutamic acids. One advantage of isolating peptides by virtue of non-terminal tags is that selection can be made for larger peptides which are more likely to contain a specific amino acid to which a tag is attached thus isolating peptides with a

mass which exceeds low molecular weight masses with a larger background noise during mass analysis.

In another aspect, the present invention provides for cleavage of proteins using proteases or chemical methods and subsequent mass analysis without further fractionation. In this case, the analysis of protein mixtures is assisted by sequential cleavage cycles whereby the spectrum of proteins and peptides are analysed following each cleavage cycle. This method could also include chemical tagging cycles between cleavage cycles to increase the mass or steps to remove side-groups such as carbohydrate groups in order to reduce mass. If the mass of the range of protein fragments is then determined at the end of each cleavage cycle (either with or without chemical tagging, cleavage or other modification), then a range of mass distributions will be obtained for each cycle. With an appropriate series of mass modification cycles, the result for a single protein or a mixture will be a mass spectrum of protein/peptide fragments which is altered at successive cycles; the pattern of these alterations will provide a "fingerprint" for the specific proteins/peptides in the mixture. The appearance and disappearance of a particular protein/peptide fragment of a certain mass following a specific cleavage cycles with or without chemical tagging, cleavage or other modifications will provide a fingerprint for identification of the fragment sequence especially by reference to a database of such fingerprints. Comparison of the spectrum of protein/peptide fragments from different related samples then allows for the identification of protein/peptide fragment differences between these samples. Particularly useful in this embodiment of the present invention is proteases which specifically recognise two amino acids and cleave the protein as a result. An example of such proteases are the prohormone convertases which cleave between dibasic amino acid pairs.

Therefore, the invention provides for novel ways of analysing protein mixtures using a combination of protein digestion or cleavage and mass analysis.

In a related aspect of the present invention, proteins are fractionated prior to cleavage. For large protein mixtures, particularly those isolated directly from whole cells or tissues, the pre-fractionation of proteins may be desirable in order to reduce the complexity of mixtures subjected to subsequent cleavage, peptide fractionation and mass analysis. Whilst affinity reagents can be used which recognise sequences or structures in the proteins/peptides directly, this will itself require a complex library of affinity reagents such as an antibody library and therefore the additional use of chemical tags to provide moieties recognised by a set of affinity reagents provides an alternative means of using such reagents. More conventional means of pre-fractionation include the use of gel electrophoresis either in one or two dimensions where sections of the gel are isolated and the proteins within then subjected to cleavage and mass analysis. Other pre-fractionation methods include isolation of proteins by virtue of natural modifications such as phosphorylation, glycosylation, protein-protein (or peptide) interaction; alternatively, membrane proteins can be pre-fractionated or proteins from particular compartments within the cell. Another important pre-fractionation procedure is to remove highly abundant proteins from the mixture using affinity reagents such as antibodies to bind and

remove such proteins. As an alternative to pre-fractionation, peptides generated after cleavage can also be fractionated by many of these means and also including size/charge fractionation methods using HPLC and by virtue of natural modifications using, for example, antibodies which bind phosphorylated amino acids within peptides. Prefractionation of proteins may also be achieved by using affinity reagents such as monoclonal/polyclonal antibodies to isolate specific proteins for subsequent cleavage and mass analysis. For such analysis of larger mixtures of proteins, panels of mixed monoclonal antibodies such as those provided by recombinant libraries of antibody variable region fragments (including single-chain antibodies) are preferred in order to isolate subsets of proteins or subsets of cleaved peptides for subsequent analysis. Such panels of monoclonal antibodies will include a wide range of protein or peptide specificities which could be achieved, for example, by pre-absorbing antibody libraries on the mixed protein/peptide sample of interest and then using individual or mixtures of the selected antibodies in order to isolate subsets of proteins or peptides. Such analysis provides mass spectra for a range of different protein/peptide fractions thus facilitating detection of differences in specific proteins between samples.

A further advantage of the use of chemical tags is that the subsequent fractionation of peptides by affinity reagents can greatly reduce the number of selected peptides from a protein molecule with the rest of the molecule thus being eliminated from the mass analysis. An especially convenient method for selective chemical tagging is to tag either (or both of) the N and C terminus of the protein molecules in the mixture and then to digest or cleave the protein molecules with a reasonably selective reagent such as a amino acid or sequence-specific protease (such as endopeptidase Arg-C) or cleavage reagent (such as acid pH to cleave at Asp-Pro). Using an affinity reagent, N or C terminal peptides (or both) from the original protein could then be isolated and all internal peptides discarded. This reduction in complexity is then sufficient for mass analysis especially using HPLC coupled to a tandem mass spectrometer to analyse the peptides *en masse* in order to identify the individual peptides from the mixture.

Alternatively, chemical tagging could be performed only after digestion/cleavage, for example with the dibasic cutters, the prohormone convertases. This would provide for tagging only at one or more internal sites of the original proteins. If the protein mixture is then subjected to a second digestion/cleavage step with a different enzyme or cleaving reagent, then the size of the tagged peptides would be reduced where a cleavage site was present in the original protein. The tagged peptides could then be fractionated using an affinity reagent and subjected to mass analysis.

In another aspect of the current invention, a protein mixture is subjected to cycles of tagging, digestion/cleavage and mass analysis, whereby mass analysis is performed only on an aliquot of the mixture resultant from use of an affinity reagent binding to the specific chemical tag and whereby the master mixture is then subjected to tagging with a different chemical tag and digestion/cleavage. This provides sequentially a range of different fragments for mass analysis. Another variation on the method involves the same initial steps as above but, having exposed new N and C termini after cleavage, one

(or both) of these new termini can then optionally be tagged with a different chemical which thus tags internal sites in the original protein. If required, the process could be repeated one or more times with a different protease or cleavage reagent, each time with the addition to the N or C terminus of a different chemical tag. In one format of the method, the whole mixture of proteins would first be tagged with two different chemical groups at each of the N and C terminus and then cleaved with a protease, such as one which specifically cuts adjacent to a specific amino acid, and tagged again at the new N and C termini with two further different chemical groups. This would result in a mixture of peptides each with chemical tags at the termini. As the N and C terminal peptides would have a specific tag, these could then be isolated from the mixture using appropriate affinity reagents. Internal peptides without either the initial N or C terminal tags could be isolated using their specific tags. The process of digestion and tagging could then be repeated to create further peptides with tags. Using specific combinations of affinity reagents for specific tags, N or C terminal or specific internal peptides from the original protein could then be isolated and selected peptides discarded to achieve a reduction in complexity sufficient for mass analysis. Where chemical tags are added to two or more amino acid side groups within peptides, sequential use of affinity tags could isolate fractions of peptides containing specific combinations of amino acids. For example, if a mixture of peptides of average length of 20 amino acids and separately tagged at lysine and phenylalanine and the mixture comprises 25% of peptides which include neither lysine or phenylalanine, 25% with lysine only, 25% with phenylalanine and 25% with both, then the separate or sequential use of specific affinity reagents either for lysine or phenylalanine will result in fractionation of peptides into four equal fractions. In practice, such a fractionation scheme will favour the binding of larger peptides to affinity reagents as these peptides are more likely to contain one or more of the specific amino acids tagged. This will bias against the very small peptides such as those with molecular weights less than 1000 daltons which, when subjected to mass spectrometry analysis, will be more likely to coincide with background noise due to fragmented peptides and other small molecules.

Where analysis of complex protein mixtures is required such as in mammalian cells or tissues, the present invention provides a main method where proteins are fractionated either before or after cleavage and the peptides are then mass analysed. The fractionation of a complex mixture of proteins or peptides either requires a correspondingly complex mixture of affinity reagents or one or more affinity reagents which can recognise features of the proteins/peptides which are the basis for fractionation. Where cleavage is conducted prior to fractionation, the most common method used in the present invention is to cleave the whole protein mixture with a protease such as trypsin or V8 (Glu-C) protease and to then selectively isolate and mass analyse certain peptides. Commonly, N or C terminal peptides (or both) from the peptide mixture are isolated typically by adding a chemical tag to the N and/or C terminus of the proteins prior to cleavage and using an affinity reagent which isolates peptides with the chemical tag. Alternatively, specific peptides (N / C terminal or otherwise) can be isolated using affinity reagents which have been selected for binding to specific peptides within specific proteins; these will then select out those peptides from the mixture for subsequent mass analysis. Selective

isolation of peptides then allows for comparative analysis of specific peptides derived from alternative protein mixtures for their relative quantities (relating to relative levels of the proteins in their respective mixtures) and, in certain cases, for modifications of the peptides.

For isolation of N or C terminal peptides, the preparation and use of affinity reagents is one important aspect of the present invention and the labelling of the N or C terminus of proteins is another important aspect. With a typical mixture of proteins from mammalian cells or tissues or from many living organisms, several of the N termini of these proteins (and some C termini) will be modified (for example, by methylation) such that addition of a chemical tag to the terminus may be blocked. In addition, a typical mixture of proteins from mammalian cells or tissues or from many living organisms, the proteins will occur at different relative levels of abundance including, commonly, certainly highly abundant proteins. Where protein mixtures from mammalian cells or tissues or from other living organisms are used for the initial selection of affinity reagents, such highly abundant proteins may dominate selection of affinity reagents and may be predominant in the final peptide mixture for mass analysis. A solution to both of these problems is to use an artificial source of mixed proteins to isolate the affinity reagents. Typically, this will be a gene expression system whereby a gene (usually cDNA) library is used to generate the proteins without N or C terminal modifications. In addition, the use of a gene expression system allows the gene library to be "normalised" to reduce or remove highly abundant genes within the library. This is typically achieved by self-annealing of the DNA (or RNA) prior to constructing the library. Therefore, a common method in the present invention is to generate proteins by expression of gene libraries (usually normalised) resulting in proteins free from significant N or C terminal modifications and, where normalised, resulting in a protein mixture free from domination by specific proteins. A typical expression system used with gene libraries is *in vitro* transcription and translation using a eukaryotic ribosome preparation; this also provides the possibility of incorporating modified amino acids into the expressed proteins. The expressed protein mixture can then be used directly for N or C terminal labelling. Other expression systems could also be used where N terminal amino groups or C terminal carboxyl groups are not modified or prevented from subsequent chemical tagging. Where modification occurs, in some cases the N terminal modification can be removed either using enzymes such as histone deacetylase or chemical methods such as limited cyanogen bromide cleavage to remove N terminal methionines. Having produced a mixture of proteins free from N/C terminal modification, chemical tags can then be added to the N/C terminal amino group(s). For the N terminus, the ϵ-amino group of lysines can be initially blocked using reagents such as citraconic anhydride or methyl acetimidate to then allow only the N terminal amino groups to react. Alternatively, the ϵ-amino group of lysines can be blocked by incorporating modified lysines into the expression system such as *in vitro* transciption / translation whereby, for example, biotin-modified lysines can be directly incorporated instead of lysines. Chemical tags can then be added selectively to the N terminus of proteins, for example using isothiocyanates of specific molecules to which an affinity reagent is available. One such example is fluorescein which is incorporated by reaction of the proteins with fluorescein

isothiocyanate allowing subsequent purification with anti-fluorescein antibodies. Alternatively, polycarboxylic chelating agents can be incorporated as isothiocyantes allowing subsequent purification with specific metals. Once the N and/or C termini of proteins in the mixture are tagged, the protein is then comprehensively and specifically cleaved either chemically or enzymatically, using proteases such as trypsin or another cleaving agent. Such cleavage thereby releases from each protein an individual tagged terminal peptide fragment, such collection of fragments which can then be purified from the mixture of untagged peptides using an appropriate affinity reagent such as an antibody specific for the chemical tag. If required, the size of the chemical tag can be increased in order to produce a larger mass for analysis; this would be useful for peptide fragments resulting from cleavage very close to the chemical tag whereby the resultant fragment might be so small as to be mass analysed within lower molecular weight "noise". The chemical tag might, for example, comprise a piece of nucleic acid attached to the peptide via a reactive group introduced during synthesis of the nucleic acid. Such a nucleic acid molecule might also be useful for isolation of the tagged peptide via annealing of the nucleic acid to a complimentary sequence.

Following chemical tagging and isolation, the recovered mixture of N/C terminal peptides are then used as a "bait" for the isolation of affinity reagents to bind to these same peptides from proteins derived directly from mammalian cells or tissues or from other living organisms. Such affinity reagents will typically derive from a library of single chain antibodies displayed as part of a particle containing the corresponding gene encoding the antibody. Examples of such particles are ribosome display particles or phage display particles, in each case where the genes from selected antibodies can be rescued in order to propagate those specific antibodies. As an alternative, large arrays of antibodies (such as recombinant single chain or Fabs, Fvs) can be screened using the N/C terminal peptide mixture and antibodies which display binding to the peptides can be recovered via the corresponding genes. As another alternative, N and/or C terminal peptides could be used to directly generate polyclonal or monoclonal antibodies by appropriate immunisation of an animal. By these means, a mixture of affinity reagents is selected which can then be used for the analysis of mixtures of proteins such as from mammalian cells or tissues or from other living organisms. Such analysis can either involve using the mixture of affinity reagents to select out N/C terminal peptides from proteins derived from mammalian cells or tissues or from other living organisms or using individual affinity reagents to select out individual peptides. The selected peptides can then be mass analysed typically by MALDI-ToF (matrix-assisted laser desorption/ionisation time-of-flight) where the individual peptides give individual charge:mass ratios which can then be used to identify the peptide amino acid constituents. MS-MS (double mass spectroscopy) peptide sequencing can subsequently be used to identify the peptide if it can be isolated. Alternatively, the new generation of Quadrupole-ToF LC-MS-MS ("Q-ToF") instruments can provide for sequential MALDI-ToF and MS-MS within the same instrument. Indeed, affinity reagents either individually or in mixtures can be immobilised either indirectly or directly onto the desorption chip inserted into the MALDI-ToF instrument and peptides can be subsequently bound via the affinity reagents on the chip. In this way, multiple peptide

fractions adsorbed by multiple affinity reagents at different loci can be analysed on a single chip. The use of recombinant proteins as the "bait" to isolate affinity reagents also provides the prospect of attaching other tags to those proteins whereby the tags are encoded by the gene sequence; for example, a C terminal polyhistidine tag (allowing subsequent purification of the tagged fragments using nickel chelates) could be incorporated, for example through PCR-mediated incorporation into the gene sequences.

The use of recombinant proteins as the "bait" to isolate affinity reagents also provides another common method of the present invention for specifically isolating peptides using tags encoded by the recombinant proteins. Such tags can be conveniently incorporated into members of the a gene (usually cDNA) library during its construction or into individual clones or groups of clones thereof using specific PCR primers encoding such tags and designed to incorporate such tags into the resultant expressed proteins. Preferably, such tags will be incorporated into the expressed proteins in all reading frames in order to produce a productively tagged protein. Such tags will preferably be incorporated via the downstream primer of a PCR reaction with the usual result that the tag is produced towards the C terminal end of the expressed protein (although upstream termination codons may prevent this in some clones). However, tags may also be incorporated at the N terminal end or in both N and C termini.

For the isolation of specific peptides from a peptide mixture, the peptide sequences can be produced synthetically (or via recombinant DNA) and then, as above, used as the "bait" to capture specific affinity reagents. These affinity reagents can then be used to isolate these same peptides from a cleaved protein mixture derived from, for example, mammalian cells or tissues or from other living organisms.

As an alternative to selectively fractionating N or C terminal peptides or specific internal peptides, modified peptides such as peptides including phosphorylated amino acids which can be isolated using antibodies which selectively bind to phosphorylated amino acids (tyrosine, threonine or serine or combinations thereof) or using immobilised Fe^{3+} to trap negatively charged peptides. Similarly, peptides modified by glycosylation and other modifications can be isolated, in some cases where the peptide modification is further derivatised in order to facilitate isolation. For example, carbohydrates can readily be modified via periodate reactions as an intermediate to adding chemical tags such as fluorescein. A particularly important aspect of the invention is the fractionation of selectively modified peptides whereby such peptides are selectively tagged by virtue of their differential exposure to tagging within the original protein environment prior to cleavage. For example, surface exposed proteins on living cells can be selectively tagged, for example with biotin, by treating the cells with a tagging agent which preferentially reacts with specific amino acid groups. An indirect method for achieving such tagging in proteins which are naturally tagged via other stimuli within cells is to apply such stimuli in order to effect tagging of the proteins. For example, receptor-associated tyrosine kinase molecules within cells can potentially be tagged (for example, phosphorylated) by addition of the receptor ligand to those cells. Following

modification, peptides are released from proteins by cleavage and then directly mass analysed or subjected to other fractionation as above prior to analysis.

Mass analysis of proteins and peptides by the present invention is preferably performed using mass spectroscopy. In particular, MALDI-ToF analysis has the capability to very accurately measure specific mass: charge ratios for individual peptides. This method has the capability for simultaneous analysis if thousands of peptides. Above 4kD, the resolution of individual peptides (and proteins) becomes poorer such that cleavage of proteins into peptide fragments is necessary in order to provide fine resolution. Recent methods of interfacing liquid chromatography separation methods (such as HPLC) with tandem mass spectroscopy has already permitted the mass spectrum analysis of protein mixtures comprising up to 200 proteins. As such proteins are analysed following protease digestion, if an average ten peptides per protein is assumed, then the method can analyse up to 2000 peptides. Using methods of the present invention whereby, for example, only tagged N terminal peptides are analysed, then up to 2000 N terminal peptides derived from up to 2000 proteins could be analysed at any one time. As this is not sensitive enough for an *en masse* analysis of mammalian proteins from cells (typically 50,000 per cell), then peptides have to be segregated into at least 25 fractions in order for these fractions all to be analysed. Such further fractionation can be achieved by the direct use of affinity reagents to label internal ends after successive protein digestion/cleavage steps following which specific affinity reagents are used to fractionate peptides according to their tags. As an alternative to standard mass spectroscopy, MALDI-ToF can be used to produce protein mass profiles which can be compared for protein mixtures from different cells.

Chemical tags are typically moieties which can be covalently attached to proteins usually at the N or C terminus. For chemical tagging of the N terminus, this is commonly undertaken at the terminal amine group. If it is necessary to avoid tagging of the ϵ-amino group of lysines, then these can be initially blocked using reagents such as citraconic anhydride or methyl acetimidate. Terminal amine groups are then reactive with a wide range of chemical reagents especially using isothiocyanates. Thereby, common antibody-recognised ligands such as dinitrophenol and fluorescein can then attach these to the N terminus for subsequent fractionation using an antibody affinity reagent. For example, the commonly used Edman reagent phenyl isothiocyanate can be used to specifically attach to the N terminus of proteins and can be derivatised if necessary with a moiety provided for subsequent binding to an affinity reagent. For chemical tagging of the C terminus, methods based on carbodiimide activation are commonly used to introduce ligands which are bound by affinity reagents. Alternatively, addition of moieties to the C terminus of proteins has been described using reverse proteolysis whereby certain proteases such as carboxypeptidase Y and lysyl endopeptidase can work in reverse to add chemical tags, commonly by way of amino acids either as derivatised amino acids with tags for binding to an affinity reagent or by way of natural sequences of amino acids which can then be specifically bound by an affinity reagent. It will be recognised that a wide range of internal amino acids can also be chemically tagged including Lys via the ϵ-amino group, Glu / Asp via the carboxyl

group, Cys via the thiol group, Ser / Thr via the hydroxyl group and Tyr via the hydroxyphenyl group. Specific derivatisations of most other amino acids have been described. It will also be recognised that post-translation protein modifications can be used for addition of chemical tags especially with glycosylation where the sugar residues are commonly oxidised by periodate to formaldehyde groups which can then react with amine-containing molecules. Other modifications which can be used to add chemical tags include lipidation, phosphorylation and metal ion addition. It will be recognised that there are a large number of methods in the art for introducing one or more chemical tags at specific sites within protein molecules or peptides.

Affinity reagents for use in the present invention are commonly monoclonal antibodies. For specific sequences or structures within proteins or peptides, a library of recombinant antibody binding sites usually in the form of Fab's, Fvs or single-chain Fv's is used where commonly the antibody binding sites are "displayed" using, for example, bacteriophage or ribosome complexes such that the gene encoding individual antibody binding sites can be recovered. For use in the present invention, libraries of antibody binding sites can be dispersed into groups, for example by picking and arraying phage plaques or picking and arraying genes in vectors for ribosome display. Such pools will usually contain antibody binding sites for several proteins or peptides such that the pools can be used for fractionation. Alternatively, the protein or peptide mixture to which libraries of antibody affinity reagents are required can be immobilised and used as the target for the pre-selection of suitable affinity reagents which are then dispersed into pools or used as individual reagents. For chemical tags, individual monoclonal antibodies are used to specifically bind to individual tags in order to achieve subsequent fractionation.

The present invention includes the use of affinity reagents other than monoclonal antibodies where such reagents can facilitate the fractionation of peptides or proteins prior to mass analysis. Such affinity reagents would include molecules of the immune which selectively bind certain peptides such as major histocompatability proteins and T cell receptors. Other affinity reagents would include protein domains commonly involved in protein-protein binding interactions such as SH1 domains. Included in the present invention is the concept of cyclising peptides including within mixtures and especially when bound to solid phases by, for example, linking cysteine residues under reducing conditions. One method for this would be to add an additional cysteine residue at an exposed N or C terminal on immobilised peptides using, for example for C terminal immobilised peptides, standard conditions of peptide synthesis or using reverse proteolysis whereby certain proteases such as carboxypeptidase Y and lysyl endopeptidase. Included in the invention is also an elegant method for further fractionating proteins or peptides by adding, usually at the N terminus, amino acids which form part of the recognition sequence of a protease which specifically cleaves at a recognition sequence of two or amino acids whereby one or more terminal amino acids in the protease recognition site is provided by the starting protein or peptide. In this manner, only a fraction of the proteins or peptides to which the new amino acids are added will be then subject to terminal protease cleavage by virtue of the newly created sequence. In this manner, proteins or peptides can be tagged with additional amino acids

usually at the N terminus creating, in a fraction of the thus tagged mixture, a specific protease cleavage site. The proteins or peptides can then, for example, be immobilised via the new terminus for example using a tagged terminal amino acid or by adding a chemical tag to the terminus, whereby an affinity reagent is then used to immobilise the tagged moieties. After removing non-immobilised untagged molecules, the proteins or peptides can then be subjected to cleavage with the specific protease which will then only cleave where the cleavage site has been generated by a combination of synthesis-derived amino acids and the original protein or peptide-derived amino acids. The cleaved peptides can then be mass analysed (or further processed prior to mass analysis) thus representing a subset of the peptide mixture. By using parallel synthesis of specific amino acids to exposed termini followed by immobilisation and cleavage, large mixtures of proteins or peptides can be fractionated on the basis of their terminal amino acid(s). An example of a protease recognition site is ile, glu, gly, arg which is cleaved between gly and arg by Factor Xa. The sequence ile, glu, gly could be synthesised onto the N terminus of a protein or peptide and thus if the adjacent amino acid in the protein or peptide sequence were arg, the cleavage site would be created and could be cleaved by Factor Xa. Other examples of protease cleavage sites are asp, asp, asp, asp, lys, cleaved by Enterokinase between asp and lys; pro, gly, ala, ala, his, tyr cleaved between his and tyr by genease I; leu, val, pro, arg, gly, ser cleaved between arg and gly by thrombin. N terminal addition of partial sequence asp, asp, asp, asp could be used to identify proteins or peptides with N terminal lys (cleaved by enterokinase), pro, gly, ala, ala, his to identify proteins/peptides with N terminal tyr (cleaved by genease), leu, val, pro, arg to identify N terminal gly, ser; or leu, val, pro, arg, gly to identify N terminal ser (cleaved by thrombin). Other proteases such as the MMP's (matrix metalloproteinases) with specific recognition sites could be used to fractionate proteins with other N terminal amino acids. Different protease recognition sites could thus be used in combination with the proteases to fractionate proteins or peptides according to the N terminal amino acid. Where proteins are used as the starting material especially from mammalian cells whereby the N terminal protein is methionine, this can be removed if required by, for example, formylation and cleavage by a bacterial protease specific for removal of terminal formylmethionine.

Affinity reagents are an important aspect of the present invention and can be used for both broad fractionation of groups of proteins/peptides or for specific fractionation of individual proteins/peptides. For fractionation, it is first necessary to prepare fractions of or individual affinity reagents which binds to a specific fraction or specific peptide and not to other fractions/peptides. A convenient method is to fractionate the proteins or peptides prior to isolation of the affinity reagents. In the case of antibodies as the affinity reagents, such proteins/peptides can then be used either to bind displayed antibodies from a library or can be used to immunise animals for generation of antisera. Where a library of recombinant antibody binding sites such as single-chain Fv's is used, gene clones encoding these can be retrieved after binding to protein/peptide fractions providing a replicable source of the affinity reagents for subsequent isolation of the specific protein/peptide fraction. Individual single-chain Fv's may, in parallel, be screened for binding specificity, for example by analysing peptide binding by MALDI-ToF. In this

case, single-chain Fv's which bind to a single peptide from a large protein mixture are retained (in practice, those binding up to three peptides are also retained) as gene clones for subsequent individual use or use within a mixture of Fv's for isolation of a protein/peptide fraction from the mixture. It will be appreciated that free N termini from proteins are often good targets for isolation of very specific antibodies and therefore capture and release of N terminal peptides from a protein will particularly favour subsequent antibody isolation. Certain Fv's may be useful for the elimination of abundant proteins or peptides from the mixture. It will be appreciated that retention and characterisation of the binding of single-chain Fv's may also provide a means to reduce redundancy by eliminating Fv's with the same specificity as other Fv's.

The various aspects of the invention cover combinations of protein digestion/cleavage and mass analysis with a preferable step of fractionation using affinity tags for specific sequences or structures in the proteins or peptides, and an optional step of chemical tagging with fractionation by virtue of these tags. The different aspects encompass different sequences of these steps as follows;

1 - repeated digestion/cleavage cycles and mass analysis
2 - digestion/cleavage, fractionation with affinity reagents, mass analysis
3 - fractionation with affinity reagents, digestion/cleavage, mass analysis
4 - terminal chemical tagging, digestion/cleavage, fractionation with tag affinity reagents, mass analysis
5 - as 3 but with additional cycle(s) of tagging, digestion/cleavage, fractionation
6 - as 4 but with repeated tagging, digestion/cleavage cycles and mass analysis

The current invention should be considered to encompass these and related protein/peptide processing steps with the core objective of reducing the complexity of protein mixtures in order to achieve mass analysis of the resultant protein/peptide fractions.

The currently common method for operation of the invention involves tagging the N and/or C terminus of a mixture of proteins (either natural or encoded by cDNA libraries), cleaving with a protease, immobilising the N and/or C terminal peptide fragments, and releasing and subjecting the peptides to mass analysis. Alternatively, the N or C termini may be modified by addition of amino acids prior to cleavage with a sequence-specific protease. Prior to mass analysis, the peptides may alternatively be used to bind antibodies whereby these antibodies have been pre-selected to fractionate the peptides or are themselves retained as affinity reagents. The mixture of proteins may be pre-fractionated, for example by size, or may be produced from cDNA libraries which are pre-fractionated by segregation of clones. The retained affinity reagents are then used to analyse complex samples of proteins whereby the antibodies are used to bind peptides which are then mass analysed.

It will be appreciated that many of the same principles described herein for the mass analysing peptides derived from natural protein populations may also be used to analyse

recombinant protein populations. One particularly favoured application in for the isolation of recombinant antibodies such as single-chain Fv's to specific target antigens especially where the antibodies are derived from human genes whereby the selected antibody may be suitable for human therapeutic or diagnostic use. In this particular application, an extensive gene library of single-chain Fv's is created from a pool of immunoglobulin cDNA's such as those derived from peripheral blood B cells in humans. If this gene library is created in such manner that a random (or semi-random) gene sequence is included within the single-chain Fv coding region, then such a random/semi-random gene sequence will generate a random/semi-random peptide sequence in individual single-chain Fv's. Such a random/semi-random gene sequence can be created using standard methods such as PCR whereby a random/semi-random synthetic oligonucleotide sequence would be used as one of a pair of primers used to amplify immunoglobulin gene fragments during the creation of the single-chain Fv gene library. If the library was created appropriately, the resultant single-chain Fv's would each include a "peptide tag" unique to that particular Fv. Preferably, the peptide tag would be C terminal to the single-chain Fv region and include, flanked between itself and the single-chain Fv region, one or more protease sensitive sites such as sites for Arg-C or Glu-C endopeptidase. If a mixture of such single-chain Fv's was produced from a suitable gene library, then this mixture could then be mixed with a target antigen (or antigens such as on cells), usually where the antigen is immobilised. This would result in specific single-chain Fv's binding to the target antigen with non-binders (or weak binders depending on the stringency of washing) being washed away. Having washed away excess antibodies, the remaining antigen/single-chain Fv complex would then be digested with the endoprotease used to cleave the introduced protease sensitive site. This would release the tagged peptide which can then be subjected to mass analysis / mass spectrometry sequencing. Having determined the sequences of tags derived from bound single-chain Fv's, corresponding synthetic oligonucleotides can then be produced and used to specifically amplify specific single-chain Fv genes from the library. These specific single-chain Fv genes can then be further used to generate corresponding single-chain Fv's which could then be retested for antigen binding either individually or as part of a small pool of isolated single-chain Fv's. Ultimately, by this method, specific single-chain Fv's can be generated with desirable antigen binding properties and, if from a human source, potential clinical utility.

It will be appreciated that many of the same principles described herein for the digestion/cleavage, fractionation and mass analysis of proteins can also be applied to other polymeric molecules such as DNA or RNA. In the case of DNA or RNA, free phosphate and hydroxyl groups at the 5' and 3' termini respectively provide a means for very specific addition of chemical tags or direct binding to a solid phase. Sequence specific restriction or modification enzymes provide for cleavage or modification of DNA molecules. Useful affinity reagents for DNA or RNA are nucleic acids themselves which can be specifically hybridised to a complimentary DNA or RNA sequence with attachment to a solid phase either before of after hybridisation. Using such methods, complex mixtures of nucleic acids can be fractionated and then subjected to mass analysis especially using mass spectrometry.

The invention is illustrated by the following examples which some not be considering as limiting in scope;

Example 1

In this example, human p53 protein was modified with a chemical tag at its N terminus, cleaved with a protease, the chemically tagged peptide then recovered using a tag-specific monoclonal antibody and the peptide then analysed by MALDI-ToF. p53 protein was a gift from Dr Borek Vojisek (University of Brno, Czech Republic). 100ug of p53 protein with the succinimide ester of (methyl sulphonyl) ethyl carbonate according to Mikolajczyk et al., *Bioconjugate Chem.*, vol 7 (1996) p150-158 in order to block lysine side-chains. The blocked protein was dissolved at 1mg/ml in 0.1M sodium bicarbonate buffer pH8.5 and NHS-SS-biotin (Pierce, Chester, UK) was added to 100ug/ml final. The reaction was carried out for 6 hours at room temperature and terminated with ethanolamine. The protein mixture was then passed down a Sephadex G25 column (Pharmacia, Milton Keynes, UK) in PBS and the void volume collected using A280 measurements of the eluates. 40ul of eluate containing 2ug p53 was then heat denatured (95c for 5 mins), cooled to 37c and 1ug endoproteinase Arg-C (from *C. histolyticum*, Calbiochem, Nottingham, UK) was added and the mixture incubated at 37c for 1 hour. Then 10ul of streptavidin-agarose (Sigma, Poole, UK) in PBS was added and the mixture shaken for 10 minutes. The agarose was pelleted at 16000g for 1 min and washed three times in TSO buffer (75mM Tris.HCl, 200mM NaCl, 0.5% N-octyl glucoside, pH8) and three times in TSMK (10mM Tris.HCl, 200mM NaCl, 5mM 2-mercaptoethanol, pH8). Finally, 10ul of a saturated solution of alpha-cyano-4-hydroxycinnamic acid in 1% aqueous trifluoroacetic acid/acetonitrile (1:1 v/v) was added to the washed beads and 1ul of this was loaded onto the mass spectrometer chip. The analysis was carried out using a Perseptive Biosystems Voyager-DE STR Biospectrometry Workstation (Perseptive Biosystems). The mass spectra were collected by adding spectra from 200 laser shots.

The results showed a major peak corresponding to the 65 amino acid N terminal Arg-C endoprotease fragment with no significant levels of other p53 Arg-C peaks.

Example 2

The method of example 1 was repeated except that the N terminal biotin-tagged peptide was used to isolate a single-chain Fv antibody fragment from a phage display library of single-chain Fv's. Subsequently, the single-chain Fv was used to isolate the N-terminal peptide fragment from a protease digest of the test protein as confirmed by MALDI-ToF. An extract of normal human brain, prepared as in example 4, was conjugated to KLH according to Harlow and Lane, "Antibodies" (1988) (Cold Spring Harbor Publications) and used to immunise two BalbC mice. 2 doses were given intra-peritoneally with an interval of 4 weeks between them. 3 to 4 days after the 2nd inoculation, the mice were sacrificed and spleens removed by dissection. Spleen mRNA preparation was then initiated using QuickPrep™ mRNA purification kit (Pharmacia) according to the manufacturer's instructions

The Pharmacia Recombinant Phage Antibody System (Pharmacia) was used to produce a library of mouse single chain Fvs (ScFv). First-strand cDNA was generated from the mRNA using M-MuLV reverse transcriptase and random hexamer primers. Antibody heavy and light chain genes were then amplified using specific heavy and light chain primers complementary to conserved sequences flanking the antibody variable domains. The 340 and 325 base pair products generated for heavy and light chain DNA respectively were separately purified following agarose gel electrophoresis. These were then assembled into a single ScFv construct using a DNA linker-primer mix to give the VH region joined by a (Gly4Ser)3 peptide to the VL region. The assembled ScFv were amplified with primers designed to insert Sfi 1 and Not 1 sites at the 5' and 3' ends respectively, giving an 800 bp product. This fragment was purified, sequentially digested with SfiI and NotI, and repurified. The fragment was then ligated into SfiI and NotI cut pCANTAB 5 phagemid vector. PCANTAB 5 contains the gene encoding the Phage Gene 3 protein (g3p) and the ScFv is inserted adjacent to the g3 signal sequence such that it will be expressed as a g3p fusion protein. Competent *E.coli* TG1 cells were transformed with the pCantab 5/ScFv phagemid then subsequently infected with the M13KO7 helper phage. The resulting recombinant phage contained DNA encoding the ScFv genes and displayed one or more copies of recombinant antibody as fusion proteins at their tips.

Phage-displayed ScFv that bind to the were then selected or enriched by panning. Briefly, the biotinylated and protease treated p53 preparation from example 1 was applied to a streptavidin-coated glass slide (Radius Biosciences, Waltham, USA) and the slide was washed four times in PBS. After blocking with 2% non-fat dry milk in PBS, the phage preparation was applied and incubated for 1 hour. After washing 10 times with TBS/0.05% Tween 20, peptide reactive recombinant phage were detected with horse radish peroxidase conjugated anti-M13 antibody and revealed with o-phenylene diamine chromogenic substrate. These phage were subsequently eluted with 0.1M glycine.HCl pH2.2 and 1mg/ml BSA and neutralised with 2M Tris base. The eluted phage were amplified in JM103 grown in 25ml J broth. Two additional rounds of panning were undertaken and finally 10 single plaques were isolated, pooled and further amplified. An aliquot of 10^{10} amplified phage was incubated for 2 hours at 4c with 0.1ug of biotinylated and endoproteinase Arg-C digested p53 in TSO buffer. After 2 hours, 0.5ug of anti-M13 (Pharmacia) in TSO was added and incubated for 1 hour following which 5ul of protein A/G agarose (Sigma) was added and the mixture incubated for a further 0.5 hours with swirling. The agarose beads were then pelleted, washed as in example 1 above and analysed by mass spectrometry.

The results showed the same major peak as in example 1 corresponding to the 65 amino acid N terminal Arg-C endoprotease fragment.

Example 3

In this example, a gene fragment encoding a test protein was subjected to priming with a synthetic oligonucleotide encoding a polyhistidine tag. The cDNAs were expressed by *in vitro* transcription and translation (IVTT) and the tagged peptide fragments were then isolated using a nickel chelate column. These fragments were then used to isolate a

single-chain Fv antibody fragment. Subsequently, the single-chain Fv was used to isolate a peptide fragment from a protease digest of the test protein as confirmed by mass spectrometry.

Example 4

The method of example 2 was repeated using a total protein preparation from cells and the chemically tagged peptide were used to isolate a collection of single-chain Fv antibody fragments. Subsequently, a mixture of twelve of these single-chain Fv's was used to isolate peptide fragments from a protease digest of the test protein and analysed by mass spectrometry.

Example 5

In this example a single-chain antibody library was produced including unique sequence signature tag. Human peripheral blood lymphocyte RNA was prepared according to standard procedures. Briefly, lymphocytes were prepared from 10ml heparinised blood taken from 16 normal healthy donors. Lymphocytes were collected following a density gradient centrifugation procedure using Lymphoprep medium (Sigma, Poole, UK). RNA was prepared using the QuickPrep system and instructions provided by the supplier (Pharmacia, St Albans, UK). Synthesis of cDNA was conducted using a cDNA synthesis kit (Pharmacia, St Albans, UK) and random hexamer primers with conditions recommended by the supplier. Immunoglobulin heavy chain variable region (Vh) and light chain variable regions (Vl) were amplified from cDNA in separate PCR mixes using primer sets designed to maximise Vh and Vl repertoires. Primer sets were as described previously (Marks J.D. et al 1991, Eur. J. Immunol. 21: 985). Vh and Vl PCR reactions were conducted using, 2.6 units of Expand™ High Fidelity PCR enzyme mix (Boehringer Mannheim, Lewes, UK.), Expand HF buffer (Boehringer), 1.5 mM $MgCl_2$, 200 μM deoxynucleotide triphosphates (dNTPs) (Life Technologies, Paisley, UK) and 25 pmoles of each primer pool. Cycles were 96°C 5 minutes, followed by [95°C 1 minute, 50°C 1 minute, 72°C 1 minute] times 5, [95°C 45 seconds, 50°C 1 minute, 72°C 1 minute 30 seconds] times 8, [95°C 45 seconds, 50°C 1 minute, 72°C 2 minutes] times 5, finishing with 72°C 5 minutes.

In a separate PCR, a linker fragment of form $(Gly_4Ser)_3$ (Huston J.S. et al 1988, *PNAS*, 85: 5879-5883) was amplified from a cloned template pSW1-ScFvD1.3 (McCafferty et al, 1990, *Nature* **348**: 522-554) using primers sets detailed previously (Marks, J. D in *Antibody Engineering*, ed Borrebaek C.A.K New York O.U.P., 1995). The 93bp linker fragment product was annealed together with an equimolar mixture of the Vh and Vl PCR products. The mixture was further amplified in a "pull through" reaction using flanking primers HuVHBACKsfi and HuFORNot as detailed in Vaughan et al (Vaughan T.J. et al 1996, *Nature Biotech.* **14**: 309-314). All fragments used in the pull-through reaction were purified free of their initial primers prior to inclusion in the reaction. Purification was conducted using the Wizard PCR Preps system from Promega (Promega, Southampton UK).

The assembled contig of form Vh-linker-Vl, was digested with restriction enzymes SfiI and NotI (Boehringer) using standard conditions and purified as above. The purified fragment was annealed with a double stranded synthetic oligonucleotide adapter mix designed to introduce a V8 protease cleavage site juxtaposed with a tract of randomised sequence in frame with the C-terminus of the Vl gene. This V8/unique sequence tag was produced by annealing a pair of synthetic oligonucleotide pools of form 5'-ggccgcgaggaagaggaa[(atg)/(can)/(agn)/(aan)/(gan)/(ttn)]$_2$gc-3' and 5'-ggccgc[(naa)/(ntc)/(ngt)/(nct)/(nag)/(cat)]$_2$ctccttctcctcgc-3'. This linker has NotI compatible ends (underlined) and therefore facilitates the insertion of the complete single chain antibody-V8/unique sequence tag fragment into SfiI-NotI prepared pCANTAB 5 (Pharmacia) phagemid vector.

The unique sequence tag was designed to avoid the introduction of stop codons and further biased to exclude encoding residues with greater than two alternative codons. By this strategy, the number of specific oligonucleotides required to identify a given de-coded peptide sequence, is minimised. In all, the unique sequence tag is able to encode 11 of the 20 amino-acids. In addition to the V8 peptidase cleavage site (a string of 4 glutamic acid residues), the sequence tag is 12 codons long. Thus from the repertoire of 11 amino acids (10 of which are encoded by either of two codons), is able to encode $11^{12}/2 = \sim 1.5 \times 10^{12}$ different peptides.

The assembled scfv fragment (Vh-linker-Vl) with SfiI and NotI prepared ends was annealed and ligated to the NotI sequence-tag adapter and re-purified. For experiments expressing the human scfv library by phage display, the complete fragment was ligated into SfiI-NotI prepared pCANTAB 5 (Pharmacia) phagemid vector, and transformed into competent TG1 E.coli.

For other experiments using *in vitro* transcription and translation (IVTT), the assembled scfv library was subcloned into SfiI NotI prepared pCANTAB5-T7. This vector is the same as the commercially available pCANTAB5 except it was modified to include the T7 promoter sequence (ttaatacgactcactata) inserted at the HindIII site at position 2235. The modification was achieved by ligation of a double-stranded synthetic DNA linker of sequence 5'- agctaatacgactcactata into HindIII cut and de-phosphorylated pCANTAB5. Recombinant clones containing the T7 promoter were selected using a diagnostic PCR.

Following ligation and transformation into competent TG1 E.coli, cells were grown for 1 hour in 1ml of SOC medium and then plated onto TYE medium with 100ug/ml ampicillin. Colonies were scraped off plates into 5ml of 2x TY broth containing ampicillin. The cultured library was used to prepare DNA for IVTT reactions.

The pCANTAB5-T7 Scfv library DNA was used in an *in vitro* translation reaction. The IVTT was conducted using the T7 Quick coupled transcription translation mix (Promega, Southampton, UK) and 10μg of the pCANTAB5-T7 Scfv library DNA in a total volume of 50μl. The translation reaction was conducted at 30°C for 90 minutes then placed on ice. In some experiments reactions were monitored for the presence of translation

products using ^{35}S-methionine incorporation assays. Reactions were stored at –70°C prior to use in binding and screening assays.

The single-chain antibody library was used to in a binding reaction to recombinant human p53 protein (Oncogene Research Products-Calbiochem, Nottingham, UK). The IVTT mix was diluted x10 fold in PBS and used in a binding assay to human recombinant p53 protein immobilised in a 96-well microplate. The p53 protein was immobilised by overnight incubation at a concentration of 100µg/ml in phosphate buffer at 4°C. The plate was washed using PBS 0.5% (w/v) BSA and the diluted IVTT mix added to the test and control wells for binding. The binding reaction was conducted at 37°C for 90 minutes. The plate was washed x3 using PBS-T (PBS + 0.05% v/v tween-20) and subjected to V8 protease digestion (Takara, Wokingham, UK). Protein fragments were collected from the supernatant and size fractionated to exclude the V8 protease and other large species before analysis by MALDI-tof.

MALDI-tof fragment analysis identified a number of peptide fragments. The peptide sequences were used to design a set of corresponding synthetic oligonucleotides. The oligonucleotides were used in a PCR based screen of the single chain library. Pfu turbo (Stratagene Europe) DNA polymerase was used to synthesise complementary strands in members of the human single-chain antibody library DNA. Following 15 rounds of thermal cycling, the product was subjected to DpnI digestion. This step depleted the mixture of parental plasmid molecules to ensure that only the newly synthesised primed products were propagated. 1µl of the reaction was transformed into TG1 competent cells and plated onto LB plates containing 100µg/ml ampicillin. Individual clones were picked, expanded and DNA prepared according to standard procedures. The DNA was used directly in a second round of screening involving IVTT, antigen binding, V8 protease digestion, MALDI-tof fragment analysis. After 2 rounds of selection 6 scFv's were isolated which bound recombinant p53.